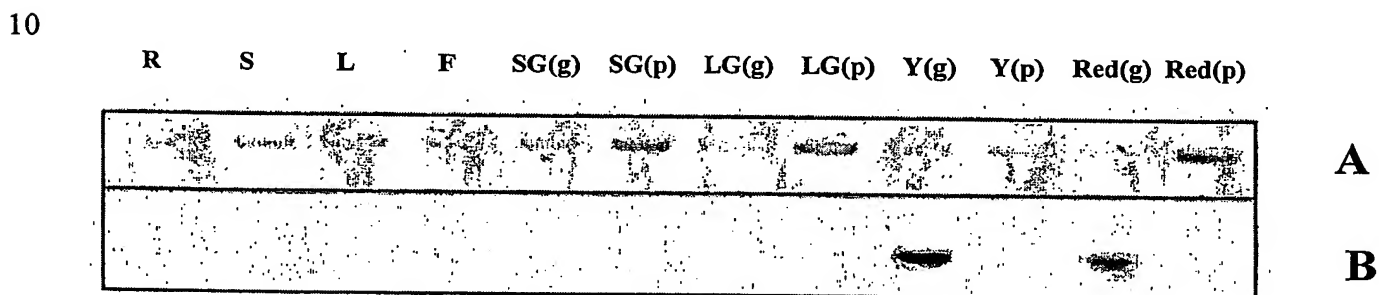
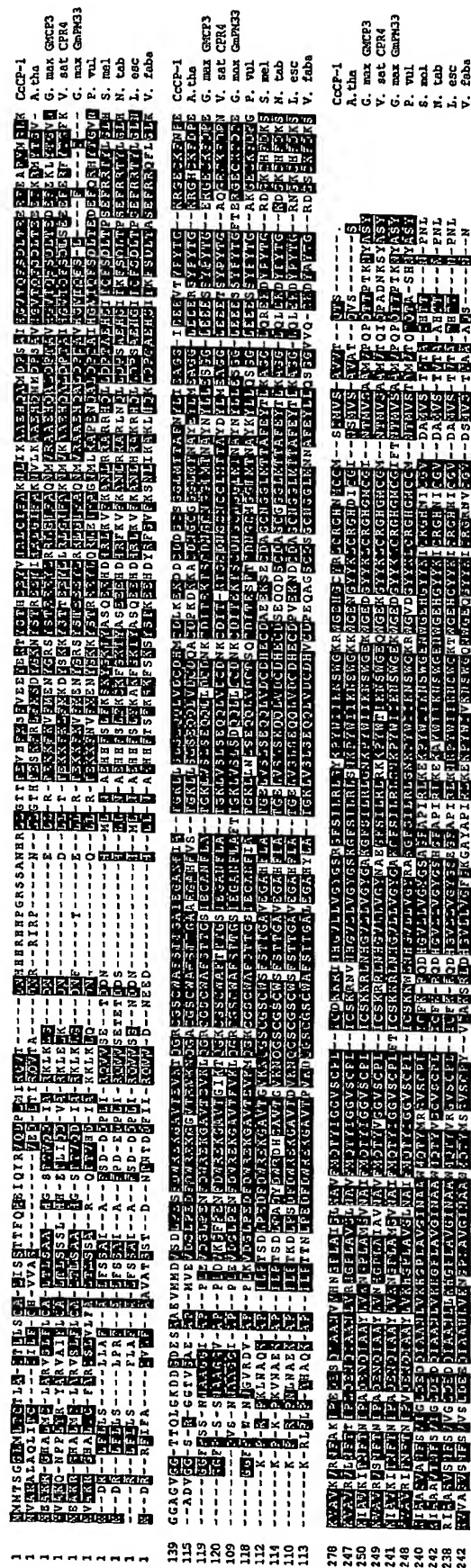


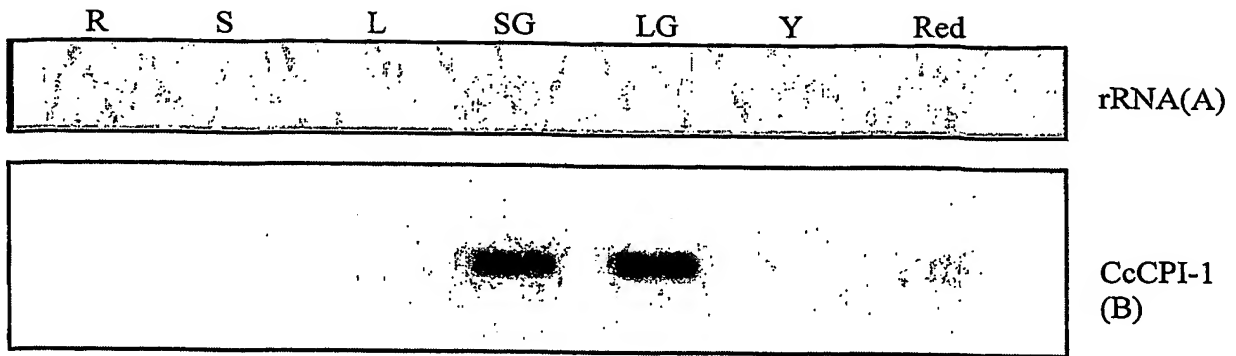
5 **Figure 1:** Northern blot analysis of the expression of the cysteine proteinase (CcCP1) gene in different tissues of *Coffea arabica*.



10
15 **Figure 2:** Northern blot analysis of the expression of the Cysteine proteinase CcCP-1 gene in different tissues of *Coffea arabica*.



5 Figure 2A: Alignment of the full sequence of the protein encoded by CcCP-1 cDNA with other full-length cysteine proteinases available in the NCBI database.



5 **Figure 3:** Northern blot analysis of the expression of the cysteine proteinase inhibitor (CcCPI-1) gene in different tissues of *Coffea arabica*.

10

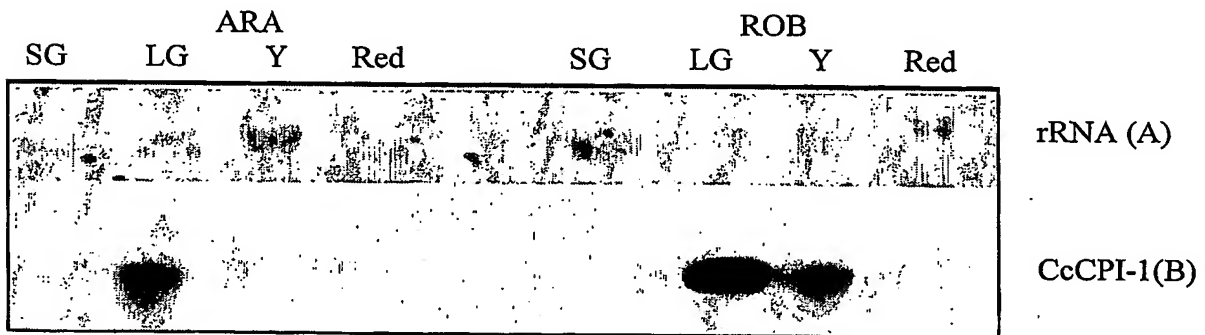


Figure 4: Northern blot analysis of the expression of the cysteine proteinase inhibitor gene (CcCPI-1) at different cherry development stages for *Coffea arabica* (ARA) and *Coffea canephora* (ROB).

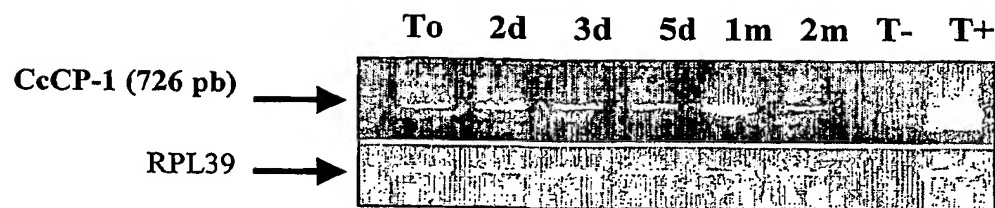


Figure 5. RT-PCR analysis of the expression of *CcCP-1* during *Coffea arabica* grain germination.

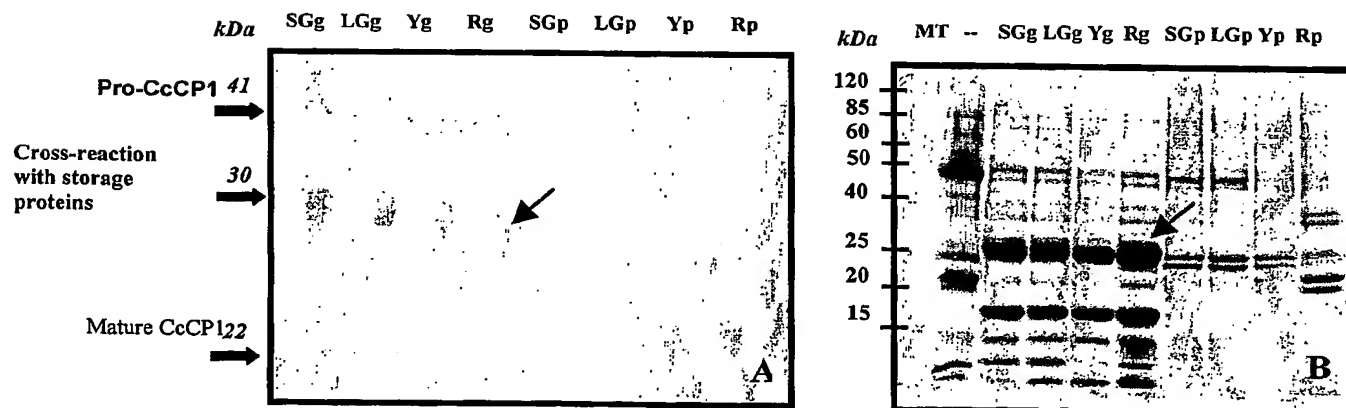


Figure 6: Western-blot analysis of the expression of *CcCP1* protein (A).

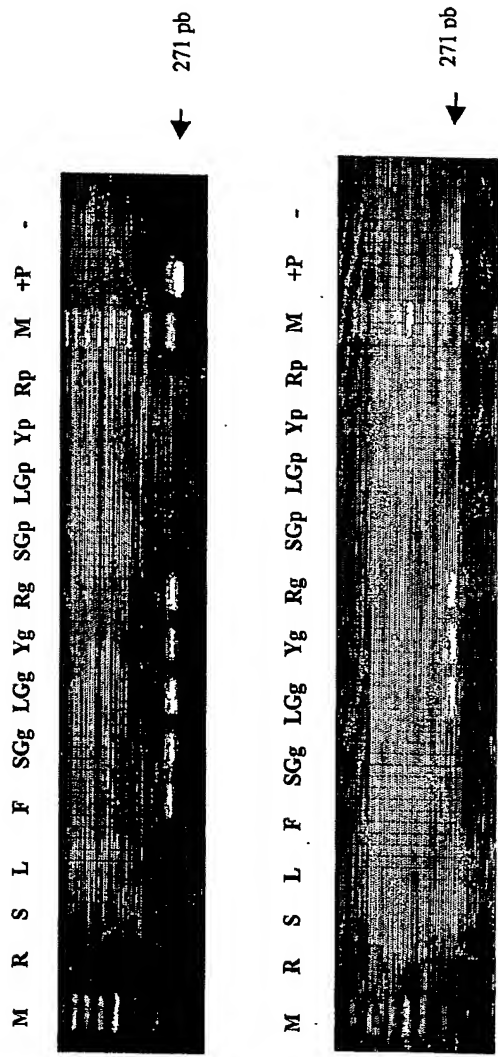


Figure 7: RT-PCR analysis of the expression of CcCPL-1 gene in different tissues of Coffea arabica CCA2 (A) and Coffea robusta FRT-32 (B).

1	M	A	K	V	G	G	I	S	E	S	K	G	N	-	E	N	S	L	E	I	E	S	L	A	K	F	A	V	D	D	Y	N	K	K	Q	N	A	L	L	E	CcCPI-2
1	M	A	T	I	G	G	I	K	Q	V	E	G	S	-	A	N	S	L	E	V	E	S	L	A	K	F	A	V	E	D	H	N	K	K	Q	N	A	M	L	E	R. obtusifolius
1	M	A	T	V	G	G	I	K	D	S	G	G	S	S	A	N	S	L	E	I	D	E	L	A	K	F	A	V	D	H	Y	N	S	K	E	N	A	L	L	E	D. caryophyllus
1	M	A	T	L	G	G	I	K	E	V	E	E	S	-	A	N	S	V	E	I	D	N	L	A	R	F	A	V	D	D	Y	N	K	K	Q	N	A	L	L	E	M. esculenta
40	F	Q	K	V	I	N	S	K	E	Q	V	V	A	G	T	V	Y	Y	L	T	I	E	V	K	D	G	N	E	K	K	L	Y	E	A	K	V	W	V	K	P	CcCPI-2
40	F	S	K	V	V	N	T	K	E	Q	V	V	A	G	T	M	Y	Y	I	T	L	E	A	T	D	G	G	K	K	K	V	Y	E	A	K	V	W	V	K	P	R. obtusifolius
41	F	Q	R	V	V	N	T	K	E	Q	V	V	A	G	T	I	Y	Y	I	T	L	E	A	T	D	G	G	V	K	K	L	Y	E	A	K	V	W	V	K	P	D. caryophyllus
40	F	K	R	V	V	S	T	K	Q	Q	V	V	A	G	T	M	Y	Y	I	T	L	E	V	A	D	G	G	Q	T	K	V	Y	E	A	K	V	W	E	K	P	M. esculenta
80	W	L	N	F	K	E	V	Q	E	F	K	P	A	A	G	D	T	S	A																						CcCPI-2
80	W	M	N	F	K	Q	V	Q	E	F	K	L	L	G	D	Q	G	S	T	S																					R. obtusifolius
81	W	V	N	F	K	E	V	Q	D	F	K	Y	V	G	D	A	S	A																						D. caryophyllus	
80	W	L	N	F	K	E	V	Q	E	F	K	P	I	G	V	A	P	S	D	S	T	A																		M. esculenta	

Figure 8: Optimal alignment of the complete protein encoded by CcCPI-2 cDNA with other homologous full-length cysteine proteinases available in the NCBI.

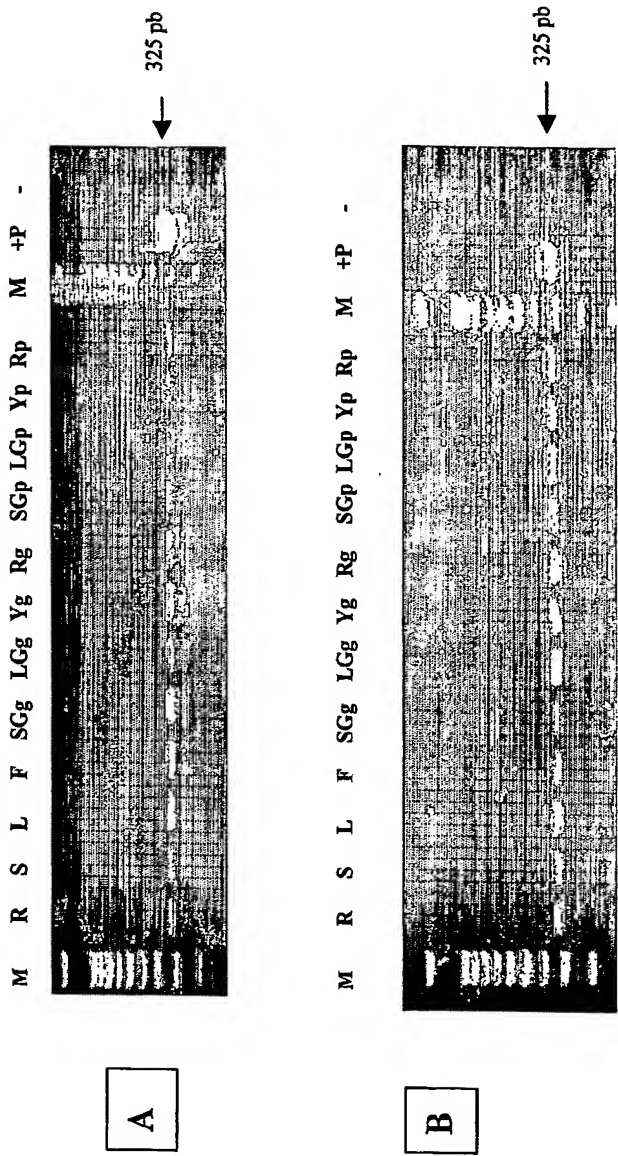


Figure 9: RT-PCR analysis of the expression of CcCPL-2 gene in different tissues of *Coffea arabica* CCA2 (A) and *Coffea robusta* FRT-32 (B).

1	M	A	S	A	F	P	H	L	L	L	L	T	T	L	A	A	I	C	L	F	S	D	V	P	S	A	A	L	G	G	R	P	K	D	A	L	V	G	G	W	CcCPI-3
1	M	N	Q	R	F	C	C	L	I	V	L	-	-	-	-	-	-	-	F	L	S	V	V	P	L	L	A	A	G	D	R	-	K	G	A	L	V	G	G	W	Citrus x paradisi
1	M	V	P	K	P	L	S	L	L	L	F	-	-	-	-	-	-	-	L	L	L	A	L	S	A	A	V	V	G	G	R	-	K	L	V	A	A	G	G	W	A. deliciosa
1	M	T	S	K	V	V	F	L	L	L	L	-	-	-	-	-	-	-	S	L	-	V	V	L	L	L	P	L	Y	A	S	-	A	A	A	R	V	G	G	W	A. thaliana
41	S	-	K	A	D	P	K	D	P	E	V	L	E	N	G	K	F	A	I	D	E	H	N	K	E	A	G	T	K	L	E	F	K	T	V	V	E	A	Q	K	CcCPI-3
33	K	P	I	E	D	P	K	E	K	H	V	M	E	I	G	Q	F	A	V	T	E	Y	N	K	Q	S	K	S	A	L	K	F	E	S	V	E	K	G	E	T	Citrus x paradisi
33	R	P	I	E	S	L	N	S	A	E	V	Q	D	V	A	Q	F	A	V	S	E	H	N	K	Q	A	N	D	E	L	Q	Y	Q	S	V	V	R	G	Y	T	A. deliciosa
32	S	P	I	S	N	V	T	D	P	Q	V	V	E	I	G	E	F	A	V	S	E	Y	N	K	R	S	E	S	G	L	K	F	E	T	V	V	S	G	E	T	A. thaliana
80	Q	V	V	A	G	T	N	Y	K	I	V	I	K	A	L	D	G	T	-	A	S	N	L	Y	E	A	I	V	W	V	K	P	W	L	K	F	K	L	T	CcCPI-3	
73	Q	V	V	S	G	T	N	Y	R	L	I	L	V	V	K	D	G	P	-	S	T	K	K	F	E	A	V	V	W	E	K	P	W	E	H	F	K	S	L	T	Citrus x paradisi
73	Q	V	V	A	G	T	N	Y	R	L	V	I	A	A	K	D	G	A	-	V	V	G	N	Y	E	A	V	V	W	D	K	P	W	M	H	F	R	N	L	T	A. deliciosa
72	Q	V	V	S	G	T	N	Y	R	L	K	V	A	A	N	D	G	D	G	V	S	K	N	Y	L	A	I	V	W	D	K	P	W	M	K	F	R	N	L	T	A. thaliana
119	S	F	R	K	L	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	CcCPI-3	
112	S	F	K	P	M	V	K	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Citrus x paradisi	
112	S	F	R	K	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A. deliciosa	
12	S	F	E	P	A	N	N	G	R	F	L	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A. thaliana

Figure 10: Optimal alignment of the complete protein encoded by CcCPI-3 cDNA with other homologous full-length cysteine proteinases available in the NCBI.

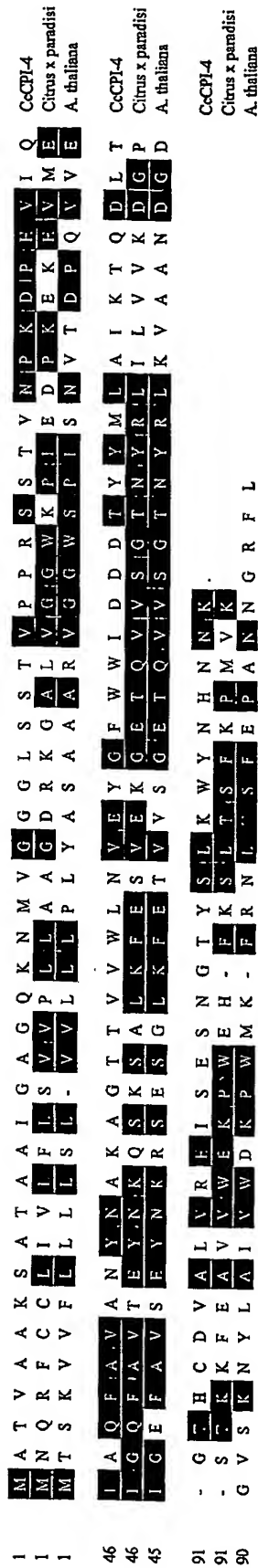


Figure 11: Optimal alignment of the complete protein encoded by CcCPI-4 cDNA with other homologous full-length cysteine proteinases available in the NCBI.

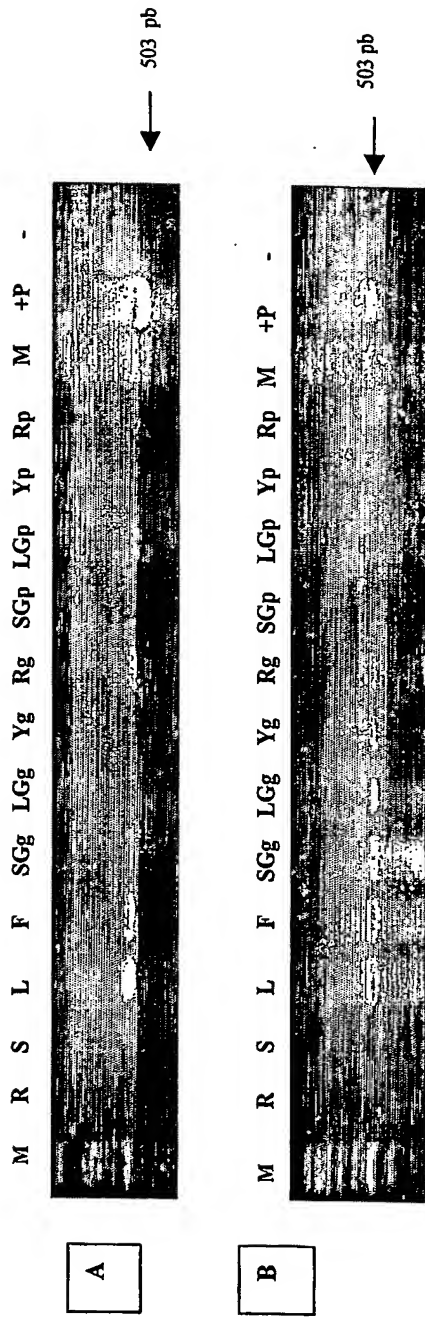


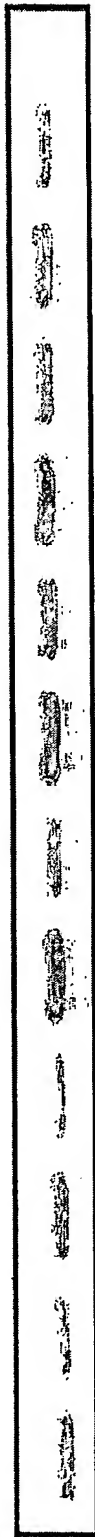
Figure 12: RT-PCR analysis of the expression of CcCPI-4 gene in different tissues of *Coffea arabica* CCA2 (Panel A) and of *Coffea robusta* FRT-32 (Panel B)

12/21

ARABICA

R S L F SG (G) SG (P) LG (G) LG (P) Y (G) Y (P) Red (G) Red (P)

tRNA (A)



CcAP-2 (B)



Figure 13: Northern blot analysis of the expression of the aspartic proteinase 2 (CcAP2) gene in different tissues of *Coffea arabica*.

```

1 gcttacatcttaaatcctgattttttatagattcgccctttcgtgaagttcaatcttcgcagtcgctcactaacatttggt
81 agacatacttcgatt ATG AAA ATG GGG AAG GCT TTC CTT TTT GCC GTT GTA TTG GCT GTG ATC
1 M K M G K A F L F A V V L A V I
144 TTA GTG GCG GCT ATG AGC ATG GAG ATC ACA GAA AGA GAT TTG GCT TCT GAG GAA AGC TTG
17 L V A A M S M E I T E R D L A S E E S L
204 TGG GAC TTG TAC GAA AGA TGG AGG AGC CAT CAT ACT GTT TCT CGA GAC CTT TCT GAG AAA
37 W D L Y E R W R S H H T V S R D L S E K
264 CGA AAG CGC TTT AAT GTT TTC AAG GCA AAT GTC CAT CAC ATT CAC AAG GTG AAC CAG AAG
57 R K R F N V F K A N V H H I H K V N Q K
324 GAC AAG CCT TAC AAG CTG AAA CTC AAC AGT TTC GCT GAT ATG ACC AAC CAC GAG TTC AGG
77 D K P Y K L K L N S F A D M T N H E F R
384 GAA TTC TAC AGT TCT AAG GTG AAA CAT TAC CGG ATG CTC CAC GGC AGT CGT GCT AAT ACT
97 E F Y S S K V K H Y R M L H G S R A N T
444 GGA TTT ATG CAT GGG AAG ACT GAA AGT TTG CCA GCC TCC GTT GAT TGG AGA AAG CAA GGA
117 G F M H G K T E S L P A S V D W R K Q G
504 GCC GTG ACT GGC GTC AAG AAT CAA GGC AAA TGT GGT AGC TGT TGG GCA TTT TCA ACT GTG
137 A V T G V K N Q G K C G S C W A F S T V
564 GTT GGA GTC GAG GGA ATC AAC AAA ATC AAA ACA GGC CAA TTA GTT TCT CTG TCC GAG CAA
157 V G V E G I N K I K T G Q L V S L S E Q
624 GAA CTT GTT GAC TGT GAA ACG GAC AAT GAA GGA TGC AAC GGA GGA CTC ATG GAA AAT GCA
177 E L V D C E T D N E G C N G G L M E N A
684 TAC GAG TTT ATT AAG AAA AGT GGG GGA ATA ACA ACT GAG AGG CTA TAT CCC TAC AAG GCA
197 Y E F I K K S G G I T T E R L Y P Y K A
744 AGA GAT GGC AGC TGT GAT TCG TCA AAG ATG AAT GCC CCT GCT GTG ACT ATT GAT GGG CAT
217 R D G S C D S S K M N A P A V T I D G H
804 GAA ATG GTA CCC GCA AAC GAT GAG AAT GCC TTG ATG AAA GCT GTT GCT AAC CAG CCT GTA
237 E M V P A N D E N A L M K A V A N Q P V
864 TCA GTA GCT ATA GAT GCG TCT GGC TCT GAC ATG CAA TTT TAT TCA GAG GGT GTA TAC GCT
257 S V A I D A S G S D M Q F Y S E G V Y A
924 GGA GAC TCG TGT GGC AAT GAG CTT GAT CAT GGC GTG GCG GTC GTC GGC TAC GGG ACT GCT
277 G D S C G N E L D H G V A V V G Y G T A
984 CTT GAC GGT ACT AAA TAC TGG ATA GTG AAG AAC TCA TGG GGA ACA GGA TGG GGA GAA CAG
297 L D G T K Y W I V K N S W G T G W G E Q
1044 GGC TAT ATC AGG ATG CAA CGT GGT GTT GAT GCT GCT GAA GGC GGA GTT TGT GGG ATA GCA
317 G Y I R M Q R G V D A A E G G V C G I A
1104 ATG GAG GCC TCC TAT CCA CTT AAA TTG TCC TCC CAC AAT CCA AAA CCA TCC CCA CCT AAG
337 M E A S Y P L K L S S H N P K P S P K
1164 GAC GAC CTC TAG attgatcctcttatatatatacatatatatatatttcagtagattcattgaatttttagttac
357 D D L *
1240 agactacgcgttcTGaagacttagatcatctctagggcatagatttatgtaaatcctgctcctgtgatggtttgaataaac
1320 aataagtagtactaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa

```

Figure 14: cDNA sequence and its deduced amino acid sequence of CcCP-4. Lowercase: 5' and 3' non-translated regions; Uppercase: Open reading frame; Bold character: amino acid sequence; *: stop codon



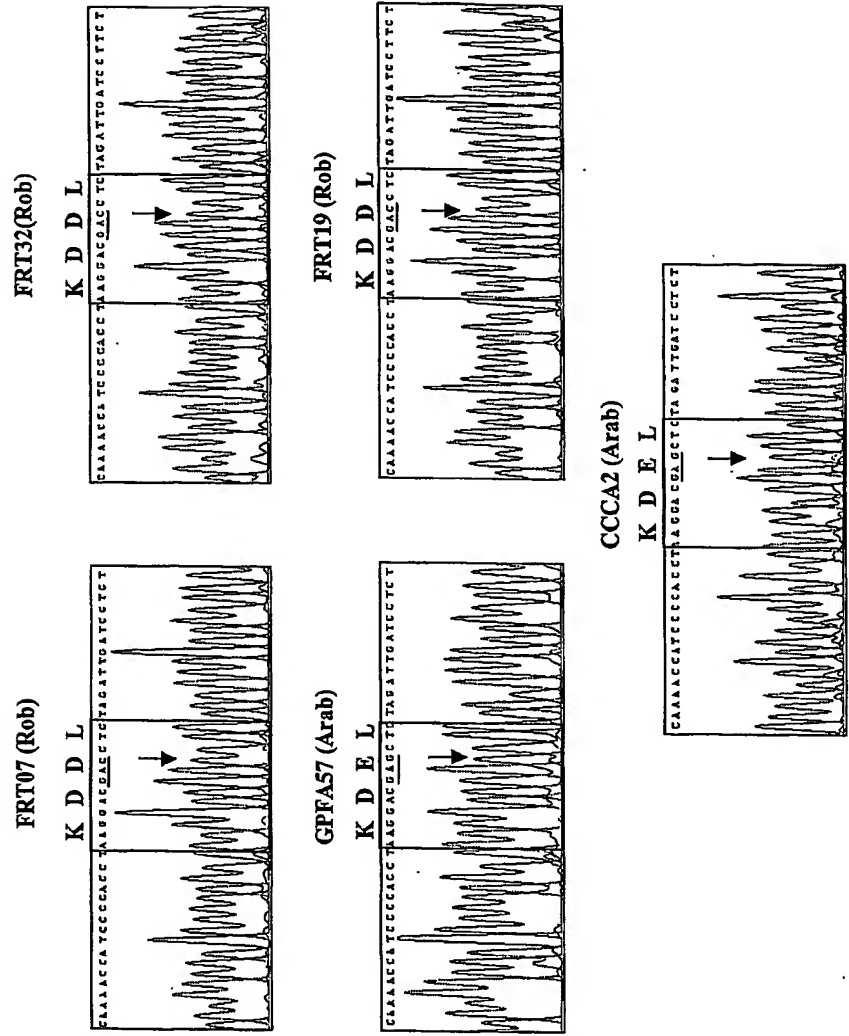
1	GGCTACATCTTAAATCCCTGATTTTATAGATTCGGCTTTTCGTGAAGTTCATCTTCCGAGTCCTCACAACATTTGCTAGACATACCTTCGATTATGA	KDEL -CCc
1	GGCTACATCTTAAATCCCTGATTTTATAGATTCGGCTTTTCGTGAAGTTCATCTTCCGAGTCCTCACAACATTTGCTAGACATACCTTCGATTATGA	KDDL -CCc
1	AAATGGGAGAGGCTTTCCTTTTTCCTGTTGATCTTACGCTGCTATGAGCAATGAGATCACAAGAGAGATTTGGCTTCTGAGAGAGAG	KDEL -CCc
101	AAATGGGAGAGGCTTTCCTTTTTCCTGTTGATCTTACGCTGCTATGAGCAATGAGATCACAAGAGAGATTTGGCTTCTGAGAGAGAG	KDDL -CCc
1	TTCTGGGACCTGCTACGAGAGATGGAGGAGGCAATCATACTGTTTTCGAGACCTTTCTGAGAGAGCAATACCGCTTTTATGCTTTCAGAGGCAATGCTCCATTC	KDEL -CCc
201	TTCTGGGACCTGCTACGAGAGATGGAGGAGGCAATCATACTGTTTTCGAGACCTTTCTGAGAGAGCAATACCGCTTTTATGCTTTCAGAGGCAATGCTCCATTC	KDDL -CCc
1	ACATTCACAGGTGAACACAGAGAGGACAGGCTTACAGCTTGAACACACACATTTTCGCTGATATGACCTAACACACAGCTTCAGGGAATTCATAGAGTTCATA	KDEL -CCc
301	ACATTCACAGGTGAACACAGAGAGGACAGGCTTACAGCTTGAACACACACATTTTCGCTGATATGACCTAACACACAGCTTCAGGGAATTCATAGAGTTCATA	KDDL -CCc
1	GGTGAACATTTACGGGATGCTCCACGGCAGTCTGCTAATACTGGATTTATGATGGGAGAGCTGAAAGTTTGGCAGGCTTCGTTGATTTGAGAGAGAGCA	KDEL -CCc
401	GGTGAACATTTACGGGATGCTCCACGGCAGTCTGCTAATACTGGATTTATGATGGGAGAGCTGAAAGTTTGGCAGGCTTCGTTGATTTGAGAGAGAGCA	KDDL -CCc
1	GGAGCCGTGACCTGGCTCAAGATCAGAGGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
501	GGAGCCGTGACCTGGCTCAAGATCAGAGGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
74	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
601	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
174	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
701	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
274	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
801	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
374	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
901	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
474	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
1001	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
574	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
1101	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
674	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
1201	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc
774	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDEL -CCc
1289	AAATAGTTTCTCTGCTGCTGAGCAAGCAATGAGTGGTGGGCAATTTTCACTGCTGGTGGAGTGGGCAATGAGTGGGCAATGAGTGGGCAATGAGTGGGCA	KDDL -CCc

Decoration 'Decoration #1': Shade (with solid black) residues that match KDDL -CCcP4 exactly.0

Figure 16. The full length cDNA sequence CcCP-4 KDDL and the partial cDNA sequence CcCP-4 (KDEL).

10

Figure 17. The complete open reading frame of CcCP-4 (KDDL) and the partial open reading frame of CcCP-4 (KDEL).



5 Figure 18. DNA sequence chromatograms for PCR amplified genomic DNA encoding the KDEI/KDDL region of the CcCP-4 gene.

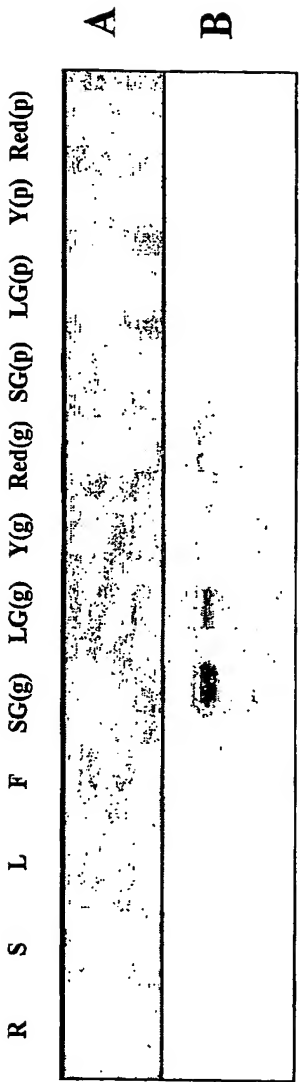


Figure 19. Northern blot analysis of the expression of the Cysteine proteinase CcCP-4 gene in different tissues of *Coffea arabica*.

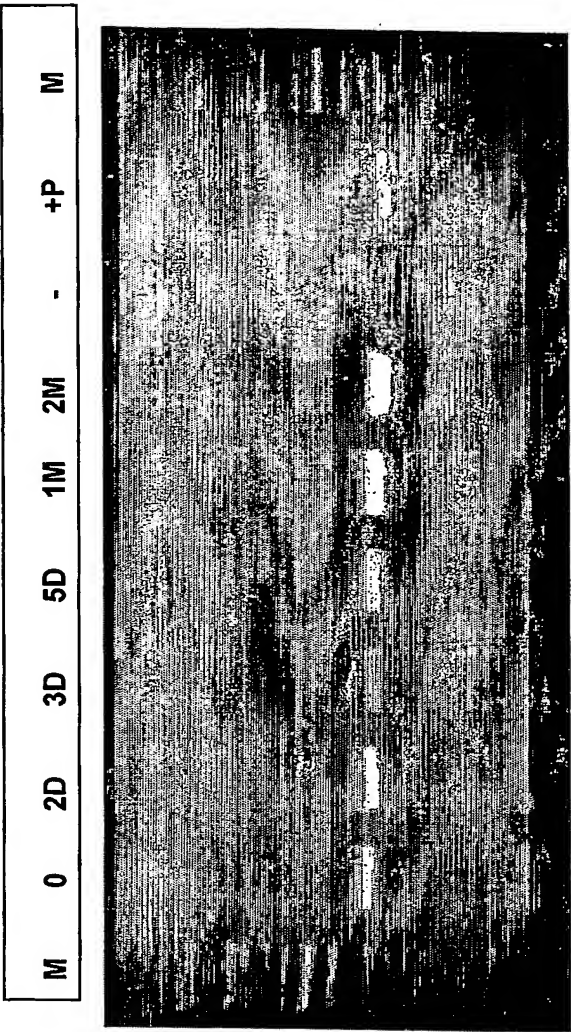


Figure 20. RT-PCR analysis of the expression of CcCP-4 in the whole grain during germination.

1 M L A L D M P L G G N G S P T D A A L V F T K L S I G CcAP1
 1 M D P S R I S R I V A V V F V L I Q V Y S G N F V F N Y T H K F A Q K E K Q L S E L K S H S D S F R H A R N L A N I D L P I G G D S R A D S I G L Y T T K I K L G APA1Ths (AY098617;
 29 T P P Q D Y Y V Q V D I G S D I I W V N C A G C V R C P K E S S L H I I D I T L V D M K A S S T G R L V T C D Q D F C L S A F N A P A S D C K V G N P C A Y S V T Y CcAP1
 1 1 S P P K E V Y V Q V D I G S D I I W V N C A P C P K C F V K I T D L G I P L S L Y D S R T S S I S K N V Q C E D D F C S - - T I M Q S E T C G A K K P C S Y H V V Y APA1Ths (AY098617;
 110 G D G S S I G G Y F V R D Y A K L M Q L F G N L Q I I P M N Q S H V F G C S S Q S S Q E L G S T E A V D G I I G F G Q A N S S I I S Q L A N A G R V K I F S H I CcAP1
 3 G D G S S I N G Y L V K D V V H I D L V T G N R O T I G S T N G T I F G C G S K Q S S Q E L G S T E A V D G I I G F G Q A N S S I I S Q L A N A G R V K I F S H I APA1Ths (AY098617;
 84 G D G S S I D G D F I K D N I T L E Q V T G N L R T I A P L A Q E V V F G C G K N Q S S Q E L G S T E A V D G I I G F G Q A N S S I I S Q L A N A G R V K I F S H I APA1Ths (BA002560)
 161 C L D G I N O G G I F A I G Q V V Q P K L K I I P L V P N E A H Y H V V L R A I E V G G D V L N L P S S D V L G G S Q S Q I I D S S G A L I L L A Y L P D D V V T F L CcAP1
 191 C L D R N N G G I F A I G E V V S P K V K T P P M L S K S A H Y S V N - N A I - V G N S V L E L S S N A F D S G D D K G V I T D S S G I T L L V Y L P D A V Y H F L APA1Ths (AY098617;
 242 C L D H M N G G I F A V G E V E S P V K I T P I V F N Q V H Y H V I E K G M D V D Q Q P I D L P P S L A S T N O D G G T I I D I S T G T T L A V L P Q N L Y H S L APA1Ths (BA002560)
 272 M E S I I A S Q S N L S I H I V E N Q F K C F V S G N V D D G F P V V X P H F E D S L S L T V A P P I E Y L F D L H D D Q W C I I G W Q N K G M Q I R D Q R E V T L CcAP1
 165 L N E I L A S H P E L T L E T V Q E S F I C F H Y I D K L D R - F F T V T F O F D K S V S L A V Y P R E Y L F Q V R E D T W C F G W O N G L Q T K G C A S L T I APA1Ths (AY098617;
 323 I E S I I A K Q - Q V K L E M V Q E T F A C F S F T S N T D K A F P V V N L E F E D S L K L S V Y P H D Y L F S L K E D M Y C F G W O S G G H T Q D C A D V I L APA1Ths (BA002560)
 353 L G D L V L A N K L V S Y D L E H Q T I G W A B Y N C S S S I K L R D E K S G N V A V G S E I I S S A R O L N A G K A L R F C L I I T S L L H A L E I P . CcAP1
 245 L G D M A L S N K L V V V D I E N V I G W T N I N C S G G I Q V E D E E S G A I V T V G A I - N L S W S S L A I T K C L I L V S L L I W V F C N V A L APA1Ths (AY098617;
 403 L G D L V L S N K L V V V D L E E N V I G W A D I H C S S S I K V E D G . S G A A V Q L G A E N L I S A S S V M N G T L V T L S I L I W V F C N V A L APA1Ths (BA002560)

Figure 21: Optimal alignment of the complete protein encoded by CcAP-1 cDNA with other homologous full-length aspartic proteinase sequences available in the NCBI.

1	MERRYLWAEFVLGAIVCSLFPLPSFG--LKKELSLKKKPLLDIQSIRAAKLAHLES THGAGRKEM--DNNL	CcAP2
1	MGQKHLLVTVCFLWALTCSLPFSFG--LKKELSLKKKPLLDIQSIRAAKLAHLES THGAGRKEM--DNNL	G.max
1	MGKRYLCNFAFLWAVVCTALPAAYSNNLRLVGLKKRPLDLDSINAAKKAAREGLRSVRPMMGAHQFI	I. batatas
1	MDKKHLCAALLLWAIACSAIPASSGD--LFRIGLKKHREIDVUSIAKAARVAKLQDRYGRHWNIGI--EKKSL	L. esculen
1	MGHRNLWVIFCFCAALISCFESTSADG--LVRIGLKKRQFSDSNSIRFVRIARAKAGMN--QGLKRF--QYSFN	N. alata
66	GSSN--EDILPLKKNYLDAQVYGEIGIGTPEQKEFTVLEFDTGSSNLWVFSKACQYFSIA CWLHHSKYYKAKKSSIS	CcAP2
67	GCKSGE--DIVVPLKKNYLDAQVYGEIGIGTPEQKEFTVLEFDTGSSNLWVFSKACQYFSIA CWLHHSKYYKAKKSSIS	G.max
65	GDSJD--EGLVSLKKNYLDAQVYGEIGIGTPEQKEFTVLEFDTGSSNLWVFSKACQYFSIA CWLHHSKYYKAKKSSIS	I. batatas
66	SDSJD--IYKVP LKKNYLDAQVYGEIGIGTPEQKEFTVLEFDTGSSNLWVFSKACQYFSIA CWLHHSKYYKAKKSSIS	L. esculen
65	GDSJD--TDIVYLKKNYLDAQVYGEIGIGTPEQKEFTVLEFDTGSSNLWVFSKACQYFSIA CWLHHSKYYKAKKSSIS	N. alata
133	TYIAIGKSCSIRYGS GSISGSFSSQDNVEVSGDLVVKQVFIELASREGSLTFVIAKFDGGL--GLGFQELIAV	CcAP2
135	THVANC TSCKINNYGTGSISGSFSSQDNVEVSGDLVVKQVFIELASREGSLTFVIAKFDGGL--GLGFQELIAV	G.max
132	TYTQIGKSCSIRYGS GSISGSFSSQDNVEVSGDLVVKQVFIELASREGSLTFVIAKFDGGL--GLGFQELIAV	I. batatas
133	TYTRDGESCSIRYGTGSISGSFSSQDNVEVSGDLVVKQVFIELASREGSLTFVIAKFDGGL--GLGFQELIAV	L. esculen
132	TYT KIGKSCFIDYGS GSISGSFSSQDNVEVSGDLVVKQVFIELASREGSLTFVIAKFDGGL--GLGFQELIAV	N. alata
201	DNMVPVWYRMV DQGLVDQVPSFSLNRPDAE DSGELVFGGVDTNHFHKGKHTYVPTQKGYWQFKMGD	CcAP2
203	ENAVPVWYRMV DQGLVDQVPSFSLNRPDAE DSGELVFGGVDTNHFHKGKHTYVPTQKGYWQFKMGD	G.max
200	ENVVPVWYDMV EQGLVDQVPSFSLNRPDAE DSGELVFGGVDTNHFHKGKHTYVPTQKGYWQFKMGD	I. batatas
201	GNTTPVWYRMV GQGLVKEPVSFSLNRPDAE DSGELVFGGVDTNHFHKGKHTYVPTQKGYWQFKMGD	L. esculen
200	GDVVPVWYRMV EQGLVSEKVPFSLNRPDAE DSGELVFGGVDTNHFHKGKHTYVPTQKGYWQFKMGD	N. alata
269	FLIGNVSTGFCGEGGCALIVDSGTSLLAGFTCVVTQINHAIGAEIGVVSIT ECKEIVSQYGEIWDITLHVSG	CcAP2
271	FLVGVSTGFCGEGGCALIVDSGTSLLAGFTCVVTQINHAIGAEIGVVSIT ECKEIVSQYGEIWDITLHVSG	G.max
268	FLIGNVSTGFCGEGGCALIVDSGTSLLAGFTCVVTQINHAIGAEIGVVSIT ECKEIVSQYGEIWDITLHVSG	I. batatas
269	FLIGNVSTGFCGEGGCALIVDSGTSLLAGFTCVVTQINHAIGAEIGVVSIT ECKEIVSQYGEIWDITLHVSG	L. esculen
268	FLIGNVSTGFCGEGGCALIVDSGTSLLAGFTCVVTQINHAIGAEIGVVSIT ECKEIVSQYGEIWDITLHVSG	N. alata
337	VLFD RVCKOAGLCLRLRGAQHENAYLKN SVVDLENKEEASVGESEFMCTACEHAYVVMQNQLKQOQSTKEKV	CcAP2
339	VKFD RVCKOAGLCLRLRGAQHENAYLKN SVVDLENKEEASVGESEFMCTACEHAYVVMQNQLKQOQSTKEKV	G.max
336	LRADQVCEGLGLCLNAGAWHES SIKTVEKEEA--EG--NLTSNPLCTCEHAYVVMQNQLKQOQSTKEKV	I. batatas
337	LRADQVCEGLGLCLNAGAWHES SIKTVEKEEA--EG--NLTSNPLCTCEHAYVVMQNQLKQOQSTKEKV	L. esculen
336	VQPUKICSQALCLCFN--DAQFLSIGIKTVIERENRKNSSVAEDFLCTACENAVVW IQNQLRREVTKEKV	N. alata
405	LAVVNQLCESIPSPNSES IIDCNLSLSTLENVVSFTIGGKSFELTLKEVVLRTGEGFAEVCISGEMPMDDV	CcAP2
406	FNVNQLCESIPSPNSES IIDCNLSLSTLENVVSFTIGGKSFELTLKEVVLRTGEGFAEVCISGEMPMDDV	G.max
402	FXYVDQLCEKLPSPNSES IIDCNLSLSTLENVVSFTIGGKSFELTLKEVVLRTGEGFAEVCISGEMPMDDV	I. batatas
404	LXYVNQLCEKLPSPNSES IIDCNLSLSTLENVVSFTIGGKSFELTLKEVVLRTGEGFAEVCISGEMPMDDV	L. esculen
403	LNYINELCDSLPSPNSESVIDCDSIPYKPNV--FTIGGKSFELTLKEVVLRTGEGFAEVCISGEMPMDDV	N. alata
473	PFERGFILWILGDFVFMGVYHTVFDYGNLRMGFAFAA	CcAP2
474	PFERGFILWILGDFVFMGVYHTVFDYGNLRMGFAFAA	G.max
470	PFERGFILWILGDFVFMGVYHTVFDYGNLRMGFAFAA	I. batatas
472	PFERGFILWILGDFVFMGVYHTVFDYGNLRMGFAFAA	L. esculen
471	PFERGFILWILGDFVFMGVYHTVFDYGNLRMGFAFAA	N. alata

Figure 22: Optimal alignment of the complete protein encoded by CcAP-2 cDNA with other homologous full-length aspartic proteinase sequences available in the NCBI